DOCUMENT RESUME

ED 179 563 TM 009 838

AUTHOR Boldt, R. F.
TITLE Some Implications of Commercial Test Normings for Mobilization Surveys.
PUB DATE Nov 78
NOTE 8p.; Paper presented at the annual meeting of the Military Testing Association (20th, Oklahoma City, OK, October 30-November 3, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bias; Data Collection; Educational Research; Motivation; *National Norms; *National Surveys; Research Design; *Research Problems; *Response Style (Tests); *Sampling; *School Involvement; Statistical Data; Statistical Surveys; Testing

ABSTRACT

Problems of obtaining a representative sample for a national survey are discussed, and the experiences of Educational Testing Service are used as examples of some methods used to cope with these problems. It is suggested that researchers: (1) capitalize on some existing program; (2) establish in the minds of the possible examinees and the possible participating schools or sampling units that they have a stake in the enterprise because of what they do; (3) sample as many sources as possible, consistent with the cost and requirements of the rest of the job; and (4) make provisions for collecting easily supplied data from everybody who was first contacted, follow up with those who don't cooperate, find out why, and modify procedures to include the nonrespondents. These procedures are suggested to be crucial in educational research, involving national surveys. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179563

TM 009838

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Some Implications of Commercial Test Normings for Mobilization Surveys

R. F. Boldt

Educational Testing Service
Princeton, New Jersey

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. F. Boldt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

November 1978

Some Implications of Commercial Test Normings for Mobilization Surveys

I assume that, in an abstract way, the object of the study contemplated by this symposium can be regarded as the description of a large segment of the American population in terms of variables that are of military interest; a major problem is that we don't normally access that segment for testing. One approach is to try to estimate the statistics of interest through their relationship with variables that are included in larger, more comprehensive surveys of populations similar to the population of interest. I will, however, focus on a different alternative--that of operating one's own survey--using the development of national test norms as an example (Jackson & Schrader, 1976).

This problem is not unlike one faced by the College Entrance Examination Board, for which we (ETS) are the technical contractors. The Board owns the Scholastic Aptitude Test (SAT), a test used for admissions in a large number of American colleges and universities. It is useful to secondary school students to know where they might stand in the SAT score distributions. To supply this, the Board periodically has us undertake to find the distribution of test scores in American secondary schools. In this, we have a situation similar to that of the military. The SAT, like the ASVAB, in whose distributional statistics we might be interested, is only given to applicants; for both testing programs there is another big pool of people out there about whom we would like to know.

Actually, the Board doesn't use the SAT to construct norms; it uses a somewhat shorter test called the Preliminary Scholastic Aptitude Test (PSAT), which is extensively administered, but at somewhat lower grades than is the SAT. Use of the PSAT is feasible for several reasons. Statistical relationships between the PSAT and the SAT have been developed, and their factorial content is highly similar, as might be expected since they are built to about the same test development specifications. Also, for the schools that appear in the sample and that also regularly administer the PSAT, the job is only one of getting them to extend their testing to include the whole class (as opposed to introducing the school to a brand new administration). Since a large portion of American secondary schools regularly administer the PSAT, the reduction of effort is very substantial. Because the tests are given at somewhat different ages, conversions that include "aging the scores" have been worked out. Indeed, the procedure of developing age relationships on smaller samples and then using the relationships to transform statistics collected on younger, but available, populations is one that might well be considered for studying variables of military interest.

Next, we choose schools as the sampling unit. It's the most feasible unit for us. ETS keeps a list of American secondary schools for the College Board; it's updated monthly and if effort makes it good, it's good. It's too big a list on which to keep detailed information about any of the schools; we have just the school name and address and, for some, the principal's name. Stratification, except geographic, directly from that list is pretty well out of the question. That such a list exists is an enormous help! With a sophisticated group like this, I hesitate to say that one of the techniques

for studying the population is to develop a frame, but I can hardly avoid saying it. The problem that I see for mobilization studies is that the population, as defined, is one for which it is very difficult to conceive a frame. But frame construction, at any rate, is one of the first things that I would undertake. Possibly, in sheer desperation, I would consider modifying the problem to be one where a frame can be constructed.

Next, the sample is to be drawn, and we must decide how many cases are necessary. Those of us who have been involved in quantitative research over the years know that much soul-searching goes with setting sample sizes. Usually, the survey sponsors want the samples to be small because that keeps the dollar cost down; the statisticians want the sample sizes up to get narrow confidence bands. For myself, I have arrived at the following attitude, one I can reach because of some statistical results about simultaneous confidence intervals. People don't ordinarily identify simultaneous confidence intervals with surveys of this kind, but I believe an important implication for us is there. We usually have several parameters that we want to estimate, and we would like to draw some conclusion about the set. Test norms comprise a whole series of statistics, and we don't just concern ourselves with the precision of each one at a time. Now, the finding that motivates my attitude about sample sizes is that if one sets many confidence intervals narrow enough to be useful, more cases are required than there is population! That's impossible, of course; it happens because the intervals are derived using infinite theory and the populations are finite. But the thrust of it is that whatever you do, it isn't going to be enough. Therefore, my attitude is that you should take what the traffic will bear; get all you can get if you're interested in multiple pieces of information. It's very seldom, in a social science inquiry, that I've encountered a situation in which any really credible confidence interval construction was involved; in fact, I can't remember <u>any</u>. Rather, there will be a certain amount of money available to do the study. What you do is to balance your resources so that you can do the best job for your sponsor and his purposes. Within that context, you get as large a sample as you can get and still get the rest of the job done--that's the state of the art. Of course, you should calculate confidence intervals for single pieces of information, and if those are too broad to be meaningful for the sample size you have, you can reasonably doubt whether the study should be performed without modification.

Therefore, by some procedure not entirely statistical, we arrived at the conclusion that there should be 200 schools! The next problem is to get their cooperation and that of their students. How (and whether) you secure that cooperation are crucial. Probably this seminar is supposed to concern technical things, but I think that careful attention to securing the cooperation of the participants is every bit as important, if not more so. In securing this cooperation, the College Board has a lot going for it. First of all, SAT scores have value, and as a consequence of participation, they can be supplied. Thus, the student has access to information that can help him forecast his position when the time comes for the grand sorting of high school students into the slots of the world, and he can get it for nothing. He gets some practice for the real thing, and the result never

hurts his record. He gets it a little bit earlier in his career than he normally would, and that helps him plan his postsecondary school career strategies. The student that really doesn't have a college career in mind and had not planned to take the PSAT may, in fact, learn that there is some desirable higher education alternative open to him that he would not have known about otherwise. (Parenthetically, since we are discussing what the examinee has to gain by taking the test, let me mention that we've had very little luck with money as an incentive for students. My belief is that the amount of money needed to buy enough student cooperation to produce a good study is more than one can afford. But I do think that there has to be an incentive for examinees.)

Schools can have advantages, too. Those not usually participating in the PSAT programs will gain experience with a national testing program with which they haven't had previous experience, and will get guidance information that they didn't have. The schools that normally have their students take the SAT or PSAT will have the national norms updated, they can use them in accustomed ways and will have more complete guidance information about their student bodies. Knowing that these things are going for them, it is perfectly reasonable for the College Board to contact American secondary schools, ask them if they would be interested and willing to participate in national norms development for the SAT, and expect to get takers. I emphasize all this strongly because I think that to conduct a national study you must approach the institutions and the examinees in such a way that they have a reason to cooperate with you. In military research you can, of course, appeal to abstractions such as patriotism, and that will be effective in some cases. But when you're pursuing cooperation of a sample, you want broad acceptance. You must have appeals with nearly universal effectiveness. Therefore I think you must somehow create an approach that establishes some gain for the participants.

In any case, a letter is written to the schools and signed by the President of the College Board. It tells about the study and tries to motivate the schools to participate. At Educational Testing Service, we're very careful about who signs such letters and how they are written. Even so, we are perhaps sometimes not as careful as we ought to be. Generally, we try to find signers who are of significance to the recipients, as do you in military testing research. (I date myself by admitting it, but I have often participated in, or observed, the drafting of a letter for the Army Adjutant General's signature.) I think the source of the letter ought to be a figure who is recognizably identified with some goal of the institution whose participation you seek; the purposes given for the study need to be purposes with real appeal to the potential participants. Recruiting is a good purpose, and evaluating the relative quality of the people coming into the service is a good purpose. But they're military purposes, and may not be as directly connected with the goals of a school principal as are other uses of his time and his students' time. I can only suggest that, and can't precisely formulate, a common interest of broad appeal be established with educational associations. I think this very important part of it needs a lot of thought.

Well, we did our best, solicited the schools, and collected the data. How well did we do? In a 1974 study, the participation rate was 58.4 of the schools. We don't know why the rate was that low. With all those things going for the study, the sample obtained still wasn't that large, so we now had the problem of deciding how much of an effect the loss of schools had on the result. One way to approach this problem is to take spot surveys of the non-respondents. If you do this, it is very possible that the results will get you to your answer. Maybe that's obvious, but let me give you an example of where it worked out very neatly. In a survey of elementary schools for the National Commission for Marihuana and Drug Abuse (Boldt, Reilly, & Haberman, 1976), we were studying the types of drug education programs that were available in elementary schools. The response rate was terrible, less than 10%. Who'd believe a 4% sample? One option was to find characteristics of that 4% sample and compare them to national characteristics and find they're the same. But that isn't very convincing as a procedure, because it doesn't explain why you got the people that you got; it merely tells you some ways that they're not different. But that doesn't establish that they're the same; it just establishes that you didn't find the ways they differ. In our case, we went back to a sample of schools who didn't respond to us, and repeatedly called them until they talked to us. Their reactions were fairly monolithic. The schools thought we had originally contacted them by mistake: What, after all, did we mean asking them about a drug abuse program in elementary school? They didn't believe there was a drug problem in elementary school and felt that the survey simply didn't apply to them. In no case was there a consciously formed drug education program. The obvious point is that anything we came out with in our 4% sample is an overestimate of the magnitude of such education. We found in our data that such education occurred in very few places, that when it did occur it occurred only because officials at the next higher levels of organization wanted it, not because of pressure from the community. The survey did establish that there wasn't local, immediate pressure because of a perceived drug problem; and there wasn't much education going on. That was part of our answer, but we got it by asking the non-respondents, not by comparing information from respondents with existing statistics. Unfortunately, many times researchers do not ask the non-participants why they didn't cooperate. With schools, sometimes, it's just that the testing area isn't big enough. If you're losing schools for a reason like that you can sample or work out some other compromise.

Another technique for the non-response problem is to send a small simple questionnaire with the original letter of request and try very hard to get it completed and returned by all persons contacted. You can then perhaps get a hypothesis about why they don't participate if, in fact, they don't. Easily supplied descriptive information would be good, and I suggest that you don't leave a place to say "we are not participating because," because frankly that makes it too easy for them not to. Make the special approach to non-participants separately.

In summary, it's best if you can capitalize on some existing program. The fact that the program exists indicates that people have an interest in

it, and if you can relate what you're doing to that interest you have a better chance of getting cooperation. Next, you need to establish in the minds of the possible examinees and the possible participating schools or sampling units, whatever they are, that they have a stake in this enterprise--not because they have a stake in what you do but because they have a stake in it because of what they do. Third, I suggest that you get all the sample that you can get, consistent with the cost and requirements of the rest of the job. Finally, make provision for collecting easily supplied data from everybody you write to in the first place, follow up those who don't cooperate, find out why, and modify your procedures, if you can, to get them in. Those aren't technical procedures (I find them remarkably unstatistical, I guess, when I think back on them), but to the extent that you need to go out and estimate national statistics, they have been crucial in our educational research.

References

Boldt, R. F., Reilly, R. R., & Haberman, P. W. A survey and assessment of drug-related problems and policies in elementary and secondary schools. In Drug use in America: Problems in perspective, Volume II of the second report of the National Commission on Marihuana and Drug Abuse, March 1973, 455-547. Also in Ronald E. Ostman (Ed.), Communication research and drug education. Beverly Hills, CA: Sage Publications, 1976. (ERIC Number ED 109 165.)

Jackson, R., & Schrader, W. B. Verbal and mathematical ability of high school juniors in 1974: A norms study of PSAT/NMSQT. RDR-76-77, No. 2. Princeton, NJ: College Entrance Examination Board, 1976.